DOCUMENT RESUME

ED 281 887 TM 870 303

AUTHOR Payne, David A.
TITLE Evaluation Realities or How I Learned to Love "The Standards" While Evaluating a Computer Assisted Instruction Project.
PUB DATE Nov 86
NOTE 13p.; Paper presented at the Annual Meeting of the American Evaluation Association (Kansas City, MO, October 29-November 1, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Case Studies; *Computer Assisted Instruction; *Evaluation Criteria; Grade 9; Grade 10; High Schools; Locus of Control; Microcomputers; Minority Groups; *Program Evaluation; School Attitudes; Self Concept; *Standards; *Student Attitudes
IDENTIFIERS *PLATO; *Standards for Evaluation Educ Prog Proj Materials

ABSTRACT

This case study presents a narrative summary of the evaluation of a two semester computer assisted instruction (CAI) project in an all minority high school. Use of PLATO software with Control Data microcomputers brought about modest achievement advantages, higher internal locus of control, more positive attitudes toward school and specific course subjects than non-CAI using ninth and tenth grade students. These effects were particularly pronounced for students in Language Skills, Chemistry, Algebra, and Mathematics. The paper concludes with discussion of lessons learned while doing the evaluation as seen in the context of the Standards for Evaluations of Educational Programs, Projects, and Materials. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED281887

ABSTRACT

Evaluation Realities or How I Learned to Love "THE STANDARDS" While Evaluating a Computer Assisted Instruction Project

David A. Payne

University of Georgia

This case study presents a narrative summary of the evaluation of a two semester CAI project in an all minority high school. Use of PLATO software with Control Data microcomputers brought about modest achievement advantages, higher internal locus of control, more positive attitudes toward school and specific course subjects than non-CAI using Ninth and Tenth grade students. These effects were particularly pronounced for students in Language Skills, Chemistry, Algebra, and Mathematics. The paper concludes with discussion of lessons learned while doing the evaluation as seen in the context of the Standards for Evaluations of Educational Programs, Projects, and Materials.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. A. Payne

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 870 303

BEST COPY AVAILABLE

EVALUATION REALITIES OR HOW I LEARNED TO LOVE "THE STANDARDS" WHILE EVALUATING A COMPUTER ASSISTED INSTRUCTION PROJECT

David A. Payne

University of Georgia

College professors are often accused of living in ivory towers away from reality. The teacher of evaluation methods is not shielded from such indictments (Payne, 1982). So it was with great expectations and in anticipation of the challenge that I accepted a contract to evaluate a computer assisted instruction program recently instituted in a local virtually all minority high school. The task was indeed challenging. Despite the practical problems of conducting an evaluation in a fast-paced and complex educational setting enthusiasm was rampant.

Following is a narrative account of that evaluation with a taste of the results. Following the summary is a discussion of what lessons were learned by doing the evaluation. In addition an attempt is made to place the evaluation in the context of the Standards for Evaluations of Educational Programs, Projects, and Materials (Joint Committee on Standards for Educational Evaluation, 1980).

Background and Setting

The use of microcomputers in the public schools is increasing every year. Use of micros ranges from tutoring, to simulation, managing, the teaching of programming, and drill and practice. Student benefits revolve around such areas as learning efficiency, learning effectiveness, pacing, convenience, and enrichment. Theoretically, the teacher should be freed to more effectively use his or her time and talents with students and eliminate redundant and

repetitious tasks. Several studies have also shown a positive impact on cognitive achievement as well as such affective educational outcomes as attitude toward school and specific school subjects, and self-concept (Kulik, Bangert, and Williams, 1983). Such outcomes were anticipated when the Atlanta Partnership of Business and Education, Inc. (a non-profit public/private change-facilitation agency) and Control Data Corporation joined forces in the installation of a computer assisted instruction program at an all minority inner city high school. The installation included 20 off-line Control Data 110 Micro-computer units, primary flexible-disk drives, and display terminals. Software came from the multi-potential PLATO (Personal Learning and Training Opportunity) population of 8,000 lessons drawn from 60 disciplines. The intent of the evaluation reported here was to assess the impact of this two semester installation.

Methodology

Because of practical limitations and administrative screw-ups, two separate data collection designs were required for the two semesters. The Fall semester phase employed a retrospective control group design (Howard, Ralph, Gulanisk, Maxwell, Nance, and Gerber, 1979). The Spring semester used a conventional post-test only control group design, with matched groups which resulted in a quasi-control group design.

Sample

Fall semester students included 346 users and 358 non-users who were primarily Ninth and Tenth graders. They averaged about 15 hours of CAI experience over the semester. Courses included Language Skills, Physics, General Math, Biology, Chemistry, Social Studies, and Algebra. Spring semester data was collected from 406 users and 126

non-users averaging 17 hours of CAI in the following courses: Survival English, General and Human Biology, Economics, Algebra, U.S. History, Political Behavior, and Math. Overlap between the two semester user/user and non-user/non-user groups was 77%.

Instrumentation

Variables measured were as follows: Academic Self Concept (9 items) (Brookover, LePere, Hamachek, Thomas, and Erickson, 1965); Locus of Control (20 items) (Crandall, Katkovsky, and Crandall, 1965); General Attitude Toward School (15 items) (Instructional Objectives Exchange, 1972); Student Opinions About Computers (19 items); Teacher Evaluation of CAI/PLATO (23 items); Attitude Toward Specific Subject (15 items) (Remmers and Silance, 1934); Final Exam scores (0-100); and Final Course Grade (A=1, ..., F=5). The General Attitude Toward School and Academic Self-Concept variables were not measured during the Spring semester as analyses of these variables First Semester data did not prove sufficiently productive or informative.

Analyses

Analysis of covariance was major statistical procedure applied to the Fall data. For the first semester a retrospective approach to data collection was applied. Students were asked to respond, for example to questions about their attitudes toward school using two frameworks. An item like, "Each morning I look forward to coming to school" was used in the context of "How I Feel Now" vs. "How I Felt at the Beginning of the School Year". Correlations between the "Now" and "Beginning" response over 15 weeks for the Self-Concept, Locus of Control, General Attitude Toward School and Attitude Toward School Subject were .74, .57, .30, and .67 respectively for the User group. For the Non-user group they were .65, .65, .42, and .59. The later

scores were used as the covariate. Spring analyses involved using a conventional analysis of variance technique.

Results

Fall semester data yielded the following results.

1. PLATO users of Language Skills, Chemistry, and Algebra had significantly higher final exam scores and end of semester grades than non-users.
2. Students and teachers evaluated PLATO software and CDC hardware very positively. Students were particularly enamored with the freedom to control the pace of their own learning (96%). Teachers preferred the clarity of objectives in the material (78%).
3. Although the statistical results were mixed, evidence suggested that students' feelings about their responsibilty for and control of their academic progress and attitudes toward specific subjects may be positively impacted by even brief exposure to CAI.

Spring semester data suggested the following conclusions.

1. No significant differences between final course grades of PLATO users and non-users. Significant final exam score differences in favor of the PLATO user group were found for students in Algebra and Mathematics.
2. Overwhelming positive evaluation of the CAI software and hardware by students. "My basic skills have been improved" (84%).
3. A significantly higher academic self-responsiblity score for PLATO users in Biology relative to non-users. This effect was significantly intensified when PLATO was used twice rather than once a week.

4. Two hours a week of PLATO has a significantly more potent positive effect on Attitude Toward School Subject than one hour a week.
5. Greater gains in internal locus of control and attitude toward school subject were observed for the user group relative to the non-user group from Fall to Spring semester.

Despite numerous methodological problems to be discussed later, it appears that the CAI had a modest positive impact on the students and to some extent on the teachers by enhancing their "computer literacy". The results are similar to those reported by Jenkins and Dankert (1981).

LESSONS FROM THE EVALUATION AND STANDARDS

The Standards for Evaluations of Educational Programs, Projects, and Materials (Joint Committee, 1980) is a set of 30 criteria grouped into four functional standards categories: Utility, Feasibility, Propriety, and Accuracy. These standards have proven useful over the years in assisting in the design of evaluation studies as well as in metaevaluation. The intent in the present section of this paper is to briefly illustrate how the standards did or could have influenced the CAI evaluation.

Utility Standards

These guidelines deal with such important issues as audience identification, evaluator credibility and several issues related to evaluation report clarity, dissemination, and timelines. In the present project there was a problem in identifying who the audiences were: Was it the hardware/software vendor, the students, the teachers, the school administrators, the board of education, or the

intermediary sponsoring facilitating agency? Although the vendor was picking up the tab for the evaluation and insisted on having considerable input into the evaluation design, it was the sponsoring agency that pulled all the pieces together. The school personnel had a vested interest because a positive evaluation might mean that the Baord would cut loose some dollars to keep the project going. As a result several reports were generated to meet the information needs of the various audiences.

The timelines issue could not be dealt with effectively. As is so often the case, particularly in educational settings, budget decisions must be made before the close of the fiscal year. As a result evaluation data, or at least a full report of such results are not available for input into the decision-making process. This was the case with present project. The Board after looking at some dollar figures, and negotiating with the vendor decided it was too expensive to continue CAI at the current level. A modest installation is, however, in operation.

Feasibility Standards

The three Feasibility standards are concerned with practicality (particularly with regard to data collection), political viability, and cost-effectiveness. Every attempt was made to minimize disruption of the ongoing program in terms of implementing the CAI program also with regard to data collection. Practicality must be balanced against accuracy of data. The evaluator in consultation with the teachers and school-level project coordinator developed, adopted or adapted instrumentation. The teachers whose students were partakers of the CAI had the responsibility for administering the inventories and working among themselves to develop the departmental achievement

tests. Having to rely on many different "data collectors" also lead to problems of incomplete data sets as motivation and investment in the project varied from teacher to teacher.

Propriety Standards

Eight standards in this category reflect on the important legal and ethical issues in conducting an evaluation. The protection of human rights and the public right to know and have access to relevant data and results are of paramount concern.

A scheme was developed so that students could code their own survey forms using a combination of the number of letters in mothers first name, an aggregate of student birthdate (e.g. 1-14-70=85), and their weight. It was felt that the gain in anonymity outweighed the potential loss in data due to inconsistent coding. Unfortunately a try out of the method proved to be a bust. The old reliable "Give me your name and I will protect it" plea was successfully used so that data could be correlated.

Periodic meetings were held with teachers and members of the Board of Education to provide progress reports and to stimulate interactions about the project particularly as regards the responsiveness of the data being gathered and reported.

Accuracy Standards

The largest number of metaevaluation criteria are to be found associated with technical adequacy. Issues related to reliability, validity, data control, and data anlaysis are distributed over eleven standards. It is frightening to perceive how inadequacies in implementation can lead to analysis problems. Following are some short-comings in program implementation which lead to less than optimal data anlaysis and interpretation.

(a) The treatment had been underway three weeks before the evaluator was consulted.

(b) Different classes used CAI for differing amounts of time, both within and across subject areas.

(c) Different classes used CAI in different ways.

(d) Non-user (control) classes were identified on the basis of school officials judgment.

(e) Students occasionally had to share terminals as there was not enough hardware or software to go around.

(f) Monitoring of student use of hardware and software was inconsistent at best.

(g) Instrumentation used was adopted from material originally developed primarily with white middle-class populations quite opposite that of the present project. Internal consistency analyses did reveal, however, acceptable levels of reliability for the students/instruments in the project. Item language was adopted as necessary.

All these threats to internal and external validity would probably cause Professors Campbell and Stanley blanch in horror. At least the evaluator was aware of them and could ameliorate overly zelous interpretations of project data and impact accordingly.

A Final Note

One of the dramatic lessons for this evaluator relates to how evaluation designs evolve. They are not simply adopted. It is an iterative process with evaluator and audience interacting at many critical points. Compromises are made as practical problems present themselves, e.g. the evaluator does not get called in until after the project is underway. Scriven calls this the "point of entry problem".

One wishes to go back up that ladder of programatic tasks to find that one last yet reversible decision in order to initiate the design process. Sometimes you can, sometimes you can't. Evaluation was also seen as evolutionary in the present project as variables were dropped from the Second semester data collection and analyses based on First semester results. The net effect was to tighten the study and reduce use of teacher and student time.

Another interesting phenomena occurs when Standards in one category interact with Standards in another to create a hypbrid problem. For example the lack of control over data collection (failure to meet a Feasibility standard) lead to a unit-of-analysis problem (failure to meet an Accuracy standard). A large amount of missing data forced abandonment of the use of the teacher as the unit-of-analysis. An individual X course unit was therefore used. As noted previously a potential achievement criterion problem was avoided by having departmental exams.

Despite competing "publics", the complex demands of practicality, and the uniqueness of the CAI experience for these students a modestly meaningful (and yes, even satisfying) evaluation was had by all. We all learned something!

REFERENCES

Brookover, W. B., LePere, J. M., Hamachek, D. E., Thomas, S., and Erickson, E. L. (1965). Self-concept of ability and school achievement: II. USDE Cooperative Research Report, Project No. 1636, East Lansing: Michigan State University.

Crandall, V. C., Katkovsky, W., and Crandall, V. J. (1965). Children's beliefs in their own control of reinforcements in intellectual-academic achievement situations. Child Development, 36, 91-109.

Howard, G. S., Ralph, K. M., Gulamick, N. A., Maxwell, S. E., Nance, D. W., and Gerber, S. K. (1979). Internal invalidity in pretest-postest self-report evaluations and a re-evaluation of retrospective pretests. Applied Psychological Measurement, 3, 1-23.

Instructional Objectives Exchange (1972). Attitude toward school. Los Angeles, CA.

Jenkins, T. M. and Dankert, E. J. (1981). Results of a three-month PLATO trial in terms of utilization and students attitudes. Educational Technology, 44-47.

Joint Committee on Standards for Educational Evaluation (1980). Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill.

Kulik, J. A., Bangert, R. L., and Williams, G. W. (1983). Effects of computer-based teaching on desondary school students. Journal of Educational Psychology, 75, (1), 19-26.

Payne, D. A. (1982). Diary of a mad evaluator. Educational Evaluation and Policy Analysis, 4, 4, 543-545.

Remmers, H. H. and Silance, E. B. (1934). Generalized attitude scales. Journal of Social Psychology, 5, 298-312.

FOOTNOTE

Tables summarizing descriptive data and analyses are available from the author: Educational Research Laboratory, UGA, Athens, GA 30602.